

# Some Reflections on the MPI Forum 1992-95

David W. Walker

Professor of High Performance Computing  
Cardiff University

<http://www.cardiff.ac.uk/people/view/118172-walker-david>



# 25 Years Ago...

- No smartphones or tablets.
- No Internet, as we know it.
- No ubiquitous wifi.
- IBM ThinkPad 700 released.
- June 1993: first Top500 list headed by the CM-5 at LANL. 1024 processors,  $R_{\max}=59.7$  Gflop/s, power=131kW.



# 25 Years Ago...

- Dow Jones opened at 3,278.69 on 24 Sept 1992.



- Bill Clinton became president.
- Trump's Plaza Hotel (New York) and two of his Atlantic City casinos were declared bankrupt in 1992.
- Hurricane Andrew hit South Florida on 22 August 1992.
- 1992 Olympics held in Barcelona.

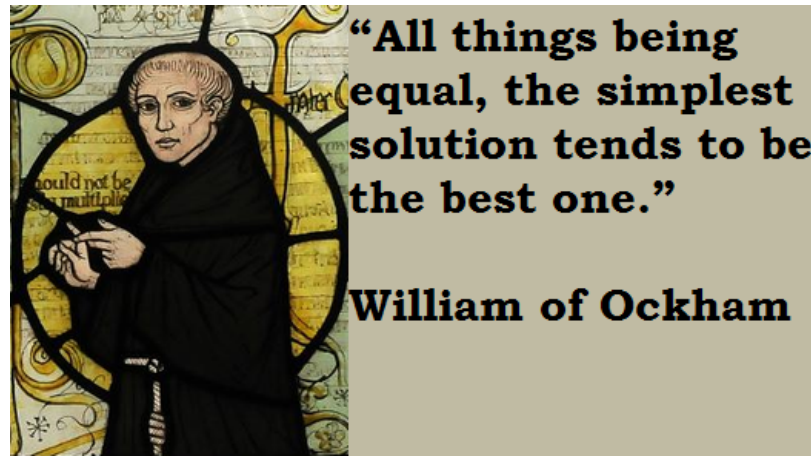
# Message Passing in the Early 1990's

- Vendors of parallel machines had their own message passing libraries, e.g., Intel's NX, CMMD on the CM-5, Vertex for the nCUBE.
- Commercial offerings, such as Express from Parasoftware.
- Portability APIs, such as P4, PARMACS, PVM, Zipcode.
- Experience with the above contributed to MPI.



# Personal Recollections

- The Bristol Suites Hotel in Dallas.
- Viewing the Perseids meteor shower from the roof of the hotel.
- Interleaving plane tickets.
- “Robust” discussions between participants.
- The role of William of Ockham.



# Written and Digital Records

- MPI-1 archive is no longer available on mpi-forum.org....
- But it is available at <http://www.netlib.org/mpi/>
- My lab note book and personal diary.
- “The Emergence of the MPI Message Passing Standard for Parallel Computing”  
[https://doi.org/10.1016/S0920-5489\(99\)00004-5](https://doi.org/10.1016/S0920-5489(99)00004-5)
- Web site: [MPI Resource Center at ORNL](#).

# Order of Events: pre-History

- **April 1992.** The Center for Research on Parallel Computation sponsored a workshop on standards for message passing in Williamsburg, VA. A summary is available. A Working Group and an email list was set up to promote work on a message passing standard.
- **Aug 1992.** Dongarra, Hempel, Hey, and Walker began work on a prototype message passing standard (MPI-0) following a meeting at a Gordon Conference in NH.
- **Nov 1992.** MPI-0 was presented at a birds-of-a-feather session at Supercomputing '92. MPI Forum was established with the aim of producing a draft message passing standard by June 1993.

# Order of Events: MPI Forum Meetings

- **January 1993.** First MPI Forum meeting held in Dallas.
- **June 1993.** After a series of five MPI Forum meetings the core of MPI, consisting of point-to-point communication routines, was completed. The minutes of the MPI Forum meetings are available.
- **November 1993.** After three more meetings of the MPI Forum the draft of the MPI specification was presented at Supercomputing '93. An overview of MPI was published in the proceedings. The MPI public comment period began.
- **January 1994.** European MPI Workshop held at INRIA, Sophia Antipolis, France.
- **March 1994.** MPI Forum meeting to tie up loose ends held in Knoxville.
- **April 1994.** End of public comment period (first comment).

# Order of Events: Publication of MPI-1

- **5 May 1994.** Final MPI specification released.
- **12 June 1994.** MPI 1.1 released. Formal end of MPI-1 standardization process.
- **July 1994.** Errata to the MPI specification published.
- **July 1997.** Revised MPI 1.2 published.

# Who Did What


- Jack Dongarra, David Walker, *Conveners and Meeting Chairs*
- Ewing Lusk, Bob Knighten, *Minutes*
- Steve Otto, Editor
- Marc Snir, William Gropp, Ewing Lusk, *Point-to-Point Communications*
- Al Geist, Marc Snir, Steve Otto, *Collective Communications*
- Rolf Hempel, *Process Topologies*
- Ewing Lusk, *Language Binding*
- William Gropp, *Environmental Management*
- James Cownie, *Profiling*
- Tony Skjellum, Lyndon Clarke, Marc Snir, Richard Littlefield, Mark Sears, *Groups, Contexts, and Communicators*
- Steven Huss-Lederman, *Implementation Issues*

# Funding and Support

- From the MPI 1.1 specification (June 1995):  
“MPI operated on a very tight budget (in reality, it had no budget when the first meeting was announced). ARPA and NSF have supported research at various institutions that have made a contribution towards travel for the U.S. academics. Support for several European participants was provided by ESPRIT.”

# ORNL Lab Book

Contains notes  
on MPI Forum  
meetings from  
25 March 1993.

Department	MSR, Bldg 6012/MS6367
Subject	GENERAL NOTES
Name	DAVID WALKER #31280
Address	OAK RIDGE NATIONAL LAB P.O. BOX 2008, BLDG. 6012 OAK RIDGE, TN 37831-6367
43-649	Phone: (615) 574-7401
 <b>Laboratory Research Notebook</b>	
<small>Dannison National Company, Holyoke, MA 01041</small>	



3/25/93

## Review of MPI draft dated 3/16/93

- Handles access opaque objects
- Opaque objects can only be created, examined, modified & destroyed by calling MPI routines, thereby hiding internal representation of MPI object
- Persistent objects ~~must~~ be explicitly destroyed. Ephemeral objects disappear once they are no longer needed.

? Does MPI- FREE block until all pending operations on object are completed?

- Process may be sequential or multithreaded
- Initial allocation of processes <sup>to</sup> processors not specified
- Process model = MIMD processes communicating via MP
- Execution model = sequential or multithreaded
- PIDs = ranks

? Does a sync before a library call prevent message conflicts? No! Consider:

Process 0	Process 1
nonblocking send (type1)	nonblocking <del>send</del> receive (type1)
non blocking send (type1)	non blocking receive (type1)
wait-send (type1)	wait-receive (type1)
wait-send (type1)	wait-receive (type1)

There are 2 ways of matching up the nonblocking sends and wait-sends in process 0, and 2 ways of matching up the nonblocking receives and wait-receives in process 1. Insertion of sync's cannot resolve ambiguity  $\Rightarrow$  must use different contexts.

- Handles and opaque objects.
- Process and execution models.
- The need for communication contexts.

## Snir - Proposal 1

Context  $\equiv$  GroupGroup operations

Copy  
Partition  
Create by list  
Deletion

Overloads

- Additional "gop" argument for pt-2-pt calls
- Sending a message in a subgroup requires table lookup.
- Context ID needed.

WHY NOT

Avoid legislating how MPI is implemented

Lyndon - Proposal VII

← Opaque object

Context is higher level concept than group.  
- group can have several contexts

Threads... should be light weight  $\Rightarrow$  don't carry state

might want to cancel  
worker when making a  
blocking communication call.

← no implicit state

Swallo agrees

- Marc Snir's "Proposal I"
- Initial ideas on groups and contexts.
- Avoid legislating how MPI is implemented.
- Thread safety seen as an issue – addressed in MPI-2

# MPI T-Shirt and RPC

To go on back of MPI T-shirt:

"You want a non-blocking what???"

Nonblocking collectives  
provided in MPI-3

MPI\_RPC ( mpi\_call, communicator, rank )

will call any MPI routine on another  
process.

# Why Was MPI-1 Successful?

- Broad support from vendors, researchers, and academics.
- US and European participants.
- Limited objectives and short time frame.
- mpich implementation available early on.
- Good dissemination through papers, books, tutorials, etc.

# Original Rationale

- **Portability and ease-of-use.** As MPI becomes more widespread it will be possible to transparently port applications between different parallel machines.
- **Provides a precise specification.** Because MPI has a formal specification hardware vendors have a well-defined set of routines that they can implement efficiently on their machines. Similarly, tool developers can build tools based on the MPI standard.

# Original Rationale

- **Necessary for growth of parallel software industry.**  
The existence of MPI makes the creation of parallel software (tools, libraries, applications, etc.) by independent software developers commercially viable. Products written using MPI or for MPI will retain their value longer and be usable on a broader machine base.
- **More widespread use of parallel computers.**  
Application developers are more likely to use parallel computers if their message-passing program is transparently portable to new and more powerful machines as they become available. Thus, the market for parallel computers will grow.

# Would the MPI-1 Effort Have Succeeded Today?

- The MPI effort is continuing.
- Less flexibility in how funding is used.
- Focus (in academia at least) is on activities that produce research papers.
- Everyone is too busy!

# Final Thoughts...

- The MPI Forum had a great camaraderie.
- People were prepared to change their ideas even if they initially took entrenched positions.
- Interactions in the MPI Forum brought the community together and led to future collaborations.
- Who would have thought that standardization could be so much fun!





Any Questions?